Molecular typing methods for outbreak detection and surveillance of invasive disease caused by *Neisseria meningitidis*, *Haemophilus influenzae* and *Streptococcus pneumoniae*, a review

Odile B. Harrison,1 Angela B. Brueggemann,1 Dominique A. Caugant,2 Arie van der Ende,3 Matthias Frosch,4 Stephen Gray,5 Sigrid Heuberger,6 Paula Krizova,7 Per Olcen,8 Mary Slack,9 Muhamed-Kheir Taha10 and Martin C. J Maiden1

Correspondence
Odile B. Harrison
odile.harrison@zoo.ox.ac.uk

1Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
2Department of Bacteriology and Immunology, Norwegian Institute of Public Health, PO Box 4404 Nydalen, NO-0403 Oslo, Norway
3Academic Medical Center, Department of Medical Microbiology, Reference Laboratory for Bacterial Meningitis, PO Box 22660, 1100 DD Amsterdam, The Netherlands
4Institut für Hygiene und Mikrobiologie, Universität Würzburg, Josef-Schneider Strasse 2, 97080 Würzburg, Germany
5Meningococcal Reference Unit, Health Protection Agency, PO Box 209, Clinical Sciences Building 2, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WZ, UK
6Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH, Bereich Humanmedizin, Institut für medizinische Mikrobiologie und Hygiene, Beethovenstraße 6, A-8010 Graz, Austria
7National Reference Laboratory for Meningococcal Infections, National Institute of Public Health, Srobarova 48, Prague, Czech Republic
8Department of Laboratory Medicine, Clinical Microbiology and Immunology, Orebro University Hospital, SE-701 85 Orebro, Sweden
9Respiratory and Systemic Infection Laboratory, Health Protection Agency Centre for Infections, 61 Colindale Avenue, London NW9 5EQ, UK
10Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France

Invasive disease caused by the encapsulated bacteria *Neisseria meningitidis*, *Haemophilus influenzae* and *Streptococcus pneumoniae* remains an important cause of morbidity and mortality worldwide, despite the introduction of successful conjugate polysaccharide vaccines that target disease-associated strains. In addition, resistance, or more accurately reduced susceptibility, to therapeutic antibiotics is spreading in populations of these organisms. There is therefore a continuing requirement for the surveillance of vaccine and non-vaccine antigens and antibiotic susceptibilities among isolates from invasive disease, which is only partially met by conventional methods. This need can be met with molecular and especially nucleotide sequence-based typing methods, which are fully developed in the case of *N. meningitidis* and which could be more widely deployed in clinical laboratories for *S. pneumoniae* and *H. influenzae*.

Introduction

The polysaccharide-encapsulated bacteria *Neisseria meningitidis* (the meningococcus), *Haemophilus influenzae* type b (Hib) and *Streptococcus pneumoniae* (the pneumococcus) are leading causes of serious bacterial infections. Together, they account for most cases of bacterial pneumonia and meningitis worldwide (Pollard et al., 2009): over 1 million children under the age of 5 years die each year from pneumococcal disease, while Hib and the meningococcus are thought to account for approximately 400 000 and 50 000 deaths annually, respectively (O’Brien et al., 2009; Scott, 2007; Tikhomirov et al., 1997; WHO, 2006). Safe and effective conjugate vaccines have been developed to provide protection against infections caused by these
organisms and offer excellent prospects for disease control; however, only 26% of children worldwide received a course of Hib vaccine in 2006 (Rossi et al., 2007), with fewer than 10% receiving other conjugate vaccines (Pollard et al., 2009).

Precise characterization of bacterial isolates from cases of invasive disease is essential for informed public health responses and the management and control of disease. Conventional typing methods involve serological approaches based on the immunochemical diversity of the capsular polysaccharide of a bacteriological isolate followed by, where available and appropriate: outermembrane protein (OMP) typing, antibiotic susceptibility profiling, PFGE fingerprinting, and/or RFLP analysis. Multi-locus enzyme electrophoresis (MLEE), the first multilocus genetic method employed for population genetic analysis of bacteria, enabled the spread of global infections by particular genotypes to be tracked, and played an important role in revealing important aspects of the population biology of these organisms (Selander et al., 1986). MLEE did not, however, achieve widespread application in routine typing of clinical isolates, due to the technical complexity of MLEE combined with difficulties in comparing results among laboratories.

During the last decade of the 20th century and the first decade of the 21st, there was a change in emphasis in microbiological typing methods, with phenotypic and serological approaches increasingly replaced by molecular techniques that indexed genotypes (Maiden & Frosch, 2001). For example, multi-locus sequence typing (MLST), a nucleotide sequence-based interpretation of MLEE, became the gold standard for N. meningitidis and S. pneumoniae isolate characterization and epidemiological surveillance, and played a major part in defining the population biology of these micro-organisms. The development of methods for the full molecular characterization of meningococci for both culture- and non-culture-confirmed cases has provided the tools necessary for enhanced surveillance and outbreak detection, and for the management of the successful implementation of major disease prevention strategies across Europe. A detailed understanding of the population biology of pneumococci has proved to be essential in identifying and understanding recent conjugate vaccine-induced changes in the pneumococcal population structure in several countries worldwide. Here, we review the nucleotide sequence-based methods currently available for these isolates, and survey existing electronic resources that enable rapid strain identification and characterization.

N. meningitidis

Epidemiology and vaccination

N. meningitidis is an accidental pathogen which colonizes around 10% of the human population in non-epidemic settings (Caugant et al., 2007). The capsular polysaccharide, which is the target for immunological serogrouping, plays an important role in virulence, enabling the bacterium to evade complement-mediated and phagocytic killing. Thirteen capsular serogroups have been described, with serogroups A, B, C, W-135 and Y responsible for over 90% of severe meningitis and septicaemia cases.

The epidemiology of meningococcal disease varies throughout the world. Most disease in Africa is caused by serogroup A meningococci, although serogroup C, W-135 and X outbreaks have been described (Greenwood, 2007). In Europe and other industrialized regions, serogroups B and C remain the major cause of invasive meningococcal disease, with serogroup B particularly prevalent in those under 20 years of age. In 2007, the overall notification rate in Europe was 1 per 100 000, with Ireland and the UK reporting the highest notification rates at 3.8 per 100 000 and 2.5 per 100 000, respectively (ECDC, 2009).

Polysaccharide vaccines against serogroup A and C meningococci are available and have been widely used, often in combination with serogroup Y and W-135 components. These vaccines, however, are poorly immunogenic in young children and infants due to their inability to mount mature anamnestic immune responses to the meningococcal capsule (Al-Mazzou et al., 2005). This has been overcome by conjugating the capsular polysaccharide to a protein carrier, which forms the basis of the meningococcal serogroup C conjugate (MCC) protein–polysaccharide conjugate vaccines that were introduced into the UK primary infant schedule in November 1999 and were available for everyone under the age of 18 years, later extended to 24 years (Girard et al., 2006). The MCC vaccines have had a major impact on disease caused by serogroup C meningococci in childhood in the UK, and have been licensed and used in other countries around Europe (Trotter & Ramsay, 2007). Currently, a combination A, C, Y and W-135 protein–polysaccharide conjugate vaccine, Menactra, is licensed for persons aged 11–55 in the USA and also in Europe, and provides effective protection against these serogroups (Pichichero et al., 2005).

Improved case ascertainment and strain characterization have been essential for accurate epidemiological surveillance. Recommendations for the molecular typing of meningococci have been developed by inter-laboratory collaboration between the European Meningococcal Reference laboratories as part of the European Monitoring Group for Meningococci (EMGM) in conjunction with the Impact of Meningococcal Epidemiology And Population Biology on Public Health in Europe (EU-MenNet) project. Many reference laboratories in Europe use the European Meningococcal Epidemiology in Real Time (EMERT) database, which since January 2007 has collected data throughout Europe in real-time (http://emrg.eu/emert/). In addition, the European Surveillance System (TESSy) will include molecular typing data for meningococcal disease (http://
Current typing approach

**N. meningitidis serogroup.** Serogroup determination can be achieved using slide agglutination assays when an isolate is available or from cerebrospinal fluid; however, early administration of antibiotic treatment hinders the ability to recover viable bacteria. Furthermore, human subjectivity in result interpretation, human error or poor quality antiserum may contribute to the misidentification of meningococcal serogroups (Mothershed et al., 2004).

Correct identification can be achieved using rapid and sensitive DNA-based methods (Mothershed et al., 2004). Furthermore, a variety of techniques have been developed that identify meningococci expressing a polysaccharide capsule that contains sialic acids, including B, C, W-135 and Y meningococci. These methods include PCR-ELISA and real-time PCR protocols (Borrow et al., 1997; Tzanakaki et al., 2003). The targeted genes encode the respective polysialyltransferases for serogroup determination. In spite of numerous publications documenting primers and protocols for the molecular identification of meningococcal serogroups, a standardized method for molecular genotyping has not yet been implemented across reference laboratories in Europe. An inter-laboratory study and quality assurance exercise, which compared genogrouping among 11 meningococcus reference centres located in 11 different countries, revealed variations in genogrouping sensitivities among all of the reference centres. These data highlight the need for further improvements and the widespread adoption of a consensus protocol (Taha et al., 2005).

**MLST.** MLST of *N. meningitidis* is based on DNA sequencing of segments of seven housekeeping loci (a♂Z, adk, aroE, fumC, gdh, pdhC, pgm), and is recognized as the gold standard for accurate strain characterization and epidemiological surveillance of this organism (Brehony et al., 2007). At the time of writing, more than 8000 sequence types (STs) and 15 000 *N. meningitidis* with complete MLST profiles had been deposited in the internet-accessible PubMLST database (http://pubmlst.org/neisseria/), submitted from diverse laboratories worldwide.

As MLST is PCR-based, characterization of *N. meningitidis* from clinical samples has now been described by several laboratories with a protocol for non-culture MLST developed by the EU-MenNet consortium and the EMGM (Diggle et al., 2003; Kriz et al., 2002). MLST data are fully accessible via the Internet (http://pubmlst.org/neisseria/), where, following interrogation of the MLST profile database, each sequence is assigned an allele number based upon its sequence.

**Fine typing: PorA and FetA**

PorA. The meningococcal class 1 OMP porin PorA is a major constituent of the outer membrane and is a serosubtyping antigen (Table 1). It is expressed by most meningococci and is a leading candidate vaccine component, although its use in this role is complicated by the diversity of this antigen among meningococci (Suker et al., 1996). The extensive variation found in PorA has led to difficulties in serosubtyping, with many variants not recognized using existing monoclonal antibody panels; however, this can be overcome using molecular techniques. Determination of the nucleotide sequence of those parts of the porA gene encoding the variable regions (VRs), VR1 and VR2, resolves immunological serosubtyping issues, enabling the peptide sequence of all variants to be deduced. This is particularly important for the development of meningococcal vaccines that include PorA. Furthermore, because of the improved typeability and discrimination achieved with porA sequence typing, the EMGM has recommended the implementation of porA sequence typing across all European reference laboratories, resulting in the development of a European protocol for porA typing based on nested PCR consensus sequences. Details of all known PorA VR variants are maintained in an online database hosted at http://pubmlst.org/neisseria/PorA/, which also provides search tools that allow sequences for both variable regions to be queried simultaneously.

FetA. The FetA protein, also known as FrpB, is an iron-regulated meningococcal OMP that has been proposed as a potential vaccine candidate (Wedge et al., 1998). The variability of the protein is similar to that of PorA, although it is limited to one VR (Thompson et al., 2003), for which various protocols exist (Table 1). The sequence diversity and length of FetA VR make it a suitable additional molecular marker, increasing the number of finetypes achieved by serogrouping and PorA sequence typing alone. At the time of writing, a total of 359 FetA VR variants had been identified (http://pubmlst.org/neisseria/FetA/).

**Antibiotic resistance profiles**

The alteration of the penA gene, encoding penicillin-binding protein 2 (PBP2), through horizontal genetic exchange has been suggested as a major mechanism for the emergence of meningococci with reduced susceptibility to penicillin, termed Pen1 isolates (Thulin et al., 2006). The modifications of PBP2 that confer the Pen1 phenotype are located in the C-terminal half of the protein, which binds penicillin and harbours the transpeptidase region (Antignac et al., 2003). Molecular approaches that detect these alterations are available, overcoming technical problems experienced between laboratories associated with MIC determination as well as culture failures (Vázquez et al., 2003).

It is currently recommended that the regions between 4948 and 5459 of the penA gene are amplified and sequenced,
Mutations in the central part of the \textit{rpoB} gene, encoding the \textit{β}-subunit of the RNA polymerase, render \textit{N. meningitidis} isolates resistant to rifampicin among \textit{N. meningitidis} isolates resistant to rifampicin (Skoczynska \textit{et al.}, 2009). Only three amino acid residues, at positions S548, H552 and S557, as well as the D542V substitution, have been found to be consistently and directly involved in conferring resistance to rifampicin among \textit{N. meningitidis} isolates, although this appears to be a rare event, perhaps as a consequence of decreased biological fitness in these organisms (Taha \textit{et al.}, 2010).

Alterations in the \textit{gyrA} and \textit{parC} genes are associated with decreased susceptibility of \textit{N. meningitidis} to quinolone antibiotics, and these variants can also be detected by PCR amplification and sequencing (Shultz \textit{et al.}, 2000). Mutations in the central part of the \textit{rpoB} gene, encoding the \textit{β}-subunit of the RNA polymerase, render \textit{N. meningitidis} isolates resistant to rifampicin (Skoczynska \textit{et al.}, 2009). Only three amino acid residues, at positions S548, H552 and S557, as well as the D542V substitution, have been found to be consistently and directly involved in conferring resistance to rifampicin among \textit{N. meningitidis} isolates, although this appears to be a rare event, perhaps as a consequence of decreased biological fitness in these organisms (Taha \textit{et al.}, 2010).

**Additional typing**

**PorB.** The meningococcal serotyping antigen PorB is a constitutively expressed OMP porin, present in most isolates. PorB proteins are divided into two classes, class 2 and class 3 meningococcal OMPs, initially based on their size as measured by SDS-PAGE, although subsequent sequence analysis has indicated substantial differences in the nucleotide and amino acid sequences of these two classes (Derrick \textit{et al.}, 1999). As with PorA serosubtyping, a range of serological reagents have been developed that detect meningococcal subtypes (Frasch \textit{et al.}, 1985); however, unlike PorA, the sequence variation in the PorB proteins that results in antigenic variation is dispersed among six of the predicted surface-exposed loops of the protein. For this reason it is recommended that PorB sequencing be undertaken in addition to PorA and FetA VR analysis only if further discrimination is required. As with PorA and FetA, an online database containing \textit{porB} alleles is available (http://pubmlst.org/neisseria/porB/), with 340 \textit{penA} alleles had been deposited at the time of writing (Table 1).

**fHbp.** The factor H binding protein (fHbp) is a surface-exposed lipoprotein present in all \textit{N. meningitidis} isolates (Fletcher \textit{et al.}, 2004). An important activity of fHbp is to bind the human complement factor H, which is thought to decrease the activation of the alternative pathway, thereby contributing to the ability of the organism to avoid complement-mediated killing by non-immune human serum or blood.

Based on sequence diversity of fHbp, different fHbp classification schemes have been proposed: the first containing three fHbp variant families, named variants 1, 2 and 3 (Massignani \textit{et al.}, 2003), the other classifying variants into subfamilies A and B (Fletcher \textit{et al.}, 2004). More recently, a unified fHbp nomenclature has been proposed, consistent with that used for other meningococcal vaccine antigens, in which unique fHbp peptide and nucleotide sequences are assigned arbitrary numeric identifiers and entered into a database similar to those for PorA, FetA and PorB (http://pubmlst.org/neisseria/fHbp/) (Brehony \textit{et al.}, 2009). This is an active database currently containing 448 new fHbp alleles.
and 378 peptide alleles at the time of writing, demonstrating the diversity of fHbp.

**Multiple-locus variable number tandem repeat analysis (MLVA)**

MLVA has been introduced as an additional typing method for a large number of bacterial pathogens, including *N. meningitidis* (Yazdankhah et al., 2005). In MLVA, the variability in the numbers of short tandem repeated sequences are indexed to create DNA fingerprints for epidemiological studies, and it has been shown to outperform MLST in discriminatory power. Eight different variable number tandem repeat (VNTR) loci with limited polymorphism for MLVA have been described and compared using a set of 85 *N. meningitidis* isolates in a Dutch study, revealing that MLVA profiles cluster similarly to groupings obtained by MLST (Schouls et al., 2006). This has application in discriminating very closely related meningococcal strains.

**H. influenzae**

**Epidemiology and vaccination**

*H. influenzae* is a Gram-negative cocco-bacillus responsible for a wide variety of respiratory infections and potentially life-threatening invasive diseases, such as meningitis and bacteraemia (Jordens & Slack, 1995). *H. influenzae* may be encapsulated (typable) with one of six polysaccharide capsules designated type a–f or unencapsulated (non-typable *H. influenzae*, NTHI). It has been estimated that at least 8 million cases of serious disease occur every year with over 370,000 deaths, the burden of disease being most significant in resource-poor countries (Watt, 2003; NTHI). It has been estimated that at least 8 million cases of serious disease occur every year with over 370,000 deaths, the burden of disease being most significant in resource-poor countries (Watt et al., 2009; WHO, 2006). The European mean incidence in those under 5 years of age in 2007 has remained stable at 0.58 per 100,000 population (ECDC, 2009).

Prior to the introduction of Hib conjugate vaccines in the early 1990s, Hib was responsible for the majority of *H. influenzae* invasive infections. The Hib vaccines currently licensed for use consist of the polyribitol phosphate (PRP) capsular polysaccharide conjugated to a protein carrier, with three types, each conjugated to a different protein carrier, currently licensed for use: PRP-OMP (OMP of *N. meningitidis*), HbOC (CRM197) and PRP-T (tetanus) (Decker & Edwards, 1998). These vaccines are safe and efficacious when administered in early infancy and are included in routine immunization programs in more than 27 European countries (ECDC, 2009). The widespread adoption of these vaccines has resulted in the near complete disappearance of serious Hib infection in children, although it does not protect against *H. influenzae* isolates that possess one of the other types of capsule (a or c–f) or against non-typable isolates. NTHi, although rarely causing life-threatening invasive infections, is a significant cause of otitis media, pneumonia and bronchitis worldwide. In addition, with the decrease in the number of Hib infections, invasive disease due to other serotypes and NTHi has increased (Campos et al., 2003), necessitating further understanding of the evolving nature of invasive *H. influenzae* infections and a comprehensive typing scheme enabling disease surveillance.

**Current typing approach**

**H. influenzae serotype.** Specific antisera that recognize the different capsule structures immunologically were developed in 1931, and serotyping by slide agglutination remains a common method of identifying the capsular type of *H. influenzae* isolates (Pittman, 1931). However, there have been concerns with respect to: (i) the overall decline in the use of and laboratory expertise in *H. influenzae* serotyping; (ii) the misinterpretation of weak or slow agglutination as a positive result; and (iii) the subjectivity in reading an agglutination reaction.

Several studies have demonstrated molecular capsule typing methods to be more sensitive and specific than slide agglutination serotyping (Falla et al., 1994; LaClaire et al., 2003; Satola et al., 2007). Encapsulated *H. influenzae* isolates contain genes for the production of their respective polysaccharide capsules at the *cap* locus, which consists of three functionally defined regions, 1, 2 and 3, for all *H. influenzae* serotypes (Kroll et al., 1991). The genes in regions 1 and 3 are common to all six capsule types, and are necessary for the processing and transport of the capsular material (*bexABCD* and *hcsAB*) (Kroll et al., 1990; Sukupolvi-Petty et al., 2006). Region 2 genes are involved in capsule biosynthesis and are unique to each of the six capsule types (Van Eldere et al., 1995). The PCR method uses multiple primer sets that recognize sequences in the Van Ketel gene (*bexA*) in region 1, at the same time as capsule-specific genes in region 2 of the *cap* locus. Comparisons between standard slide agglutination serotyping and PCR capsule typing have favoured the latter, with the most common error being the misidentification of NTHi isolates as being encapsulated (in most cases mistakenly typed as serotype d) by slide agglutination (LaClaire et al., 2003; Satola et al., 2007).

Eight biotypes, I–VIII, based on biochemical reactions that detect the production of indole, urease and ornithine decarboxylase, are also used to subdivide *H. influenzae* isolates (Kilian et al., 1979). This method uses standard biochemical tests and is still routinely employed.

**MLST.** An MLST typing scheme based on assigning allele numbers to partial sequences of seven housekeeping enzyme genes (*adk, atpG, frdB, fucK, mdh, pg, recA*) has been developed for the characterization of capsulated and unencapsulated *H. influenzae* isolates (Meats et al., 2003). Using this molecular approach, *H. influenzae* isolates with different capsules as well as unencapsulated isolates can be distinguished. For example, studies of invasive *H. influenzae* infections in Canada, in which isolates...
collected from 2000 to 2006 were typed by MLST, revealed population structuring by serotype and ST (Sill et al., 2007), while MLST typing of NTHI H. influenzae isolates showed that most isolates were members of distinct phylogenetic groups (Erwin et al., 2008). An H. influenzae MLST website is publicly available at http://haemophilus.mlst.net, in which sequence types can be assigned (Table 2). At present, nearly 900 STs and over 1600 isolates have been submitted to the MLST website, many of which detail the ST, serotype, country of origin and site from which the isolate was obtained, i.e. middle ear, sputum, etc. However, some H. influenzae isolates may be untypable using the MLST scheme due to a complete deletion of the fucose operon in these isolates (Ridderberg et al., 2010), and some modification of this scheme may therefore be required.

Antibiotic susceptibility profiles. Resistance to ampicillin and other β-lactam antibiotics is usually due to the production of a plasmid-encoded β-lactamase, TEM-1 or ROB-1 (Scriber et al., 1994). More recently, a large conjugative plasmid, recognized to be part of the Haemophilus integrating and conjugative elements (ICEs), was described among Hib isolates from Europe, and correlated with resistance to ampicillin, chloramphenicol and tetracycline (Leaves et al., 2000). Detection of ICE-specific sequences has been developed using multiplex PCR (Saha et al., 2008b).

The most important β-lactam resistance mechanism in β-lactamase-negative ampicillin-resistant isolates (termed BLNAR isolates) is the alteration of PBP3, leading to decreased affinity for penicillins and cephalosporins as a consequence of sequence variations in the gene encoding PBP3, fsl, resulting in amino acid substitutions in the transpeptidase domain of PBP3 (Parr & Bryan, 1984; Uebukata et al., 2001). A genetically based classification system for H. influenzae isolates with PBP3-mediated β-lactam resistance has been described (Dabernat et al., 2002; Hasegawa et al., 2004; Tristram & Burdach, 2007; Uebukata et al., 2001); however, an accessible online database linking isolate epidemiological data to β-lactam classification is not currently available (Table 2). Chloramphenicol resistance is usually mediated by a chloramphenicol acetyltransferase, encoded by the cat gene (Schwarz et al., 2004).

Potential fine typing candidates: P2 and P5, HMW1 and HMW2

P2 and P5. The OMPs P2 and P5 have properties comparable with those of the meningococcal PorA and FetA OMPs, and sequence variability within P2 and P5 could be used for H. influenzae in a similar way. The OMPs P2 and P5 are present in both capsulated and unencapsulated H. influenzae isolates as well as in the related swine pathogen Haemophilus parasuis (Mullins et al., 2009) (Table 2).

P2, encoded by the ompP2 gene, is an immunodominant porin with considerable antigenic heterogeneity among isolates (Sikkema & Murphy, 1992), and has been targeted as a potential vaccine candidate (Hiltke et al., 2002). P2 may play a role in colonization and has been shown to bind to specific components of human mucin. OmpP2 is additionally included during PCR capsule typing of H. influenzae as a marker for the organism. The P5 protein, encoded by the ompP5 gene, is a member of the OmpA family of OMPs, which are present as a major structural protein in many Gram-negative bacteria (Davies & Lee, 2004). P5 has both immunodominant and host-adhesive domains, and, similarly to P2, has been shown to bind to human mucin, as well as to surface-expressed carcinoembryonic antigen-related cell adhesion molecule 1 (Hill et al., 2001). Structure predictions for numerous P5 sequence variants suggest the presence of eight transmembrane regions and four surface-exposed loop regions (Webb & Cripps, 1998).

HMW1 and HMW2. The surface-exposed adhesion proteins HMW1 (high molecular weight 1) and HMW2 are potential targets for H. influenzae typing. These genes may be found among both encapsulated and unencapsulated H. influenzae isolates (Table 2), encoded by separate chromosomal loci, hmw1 and hmw2, respectively. Each locus contains an hmwa gene, which encodes the structural protein, and two accessory genes, called hmwb and hmwc, encoding proteins involved in the processing and surface localization of the HMW adhesins (Grass & St Geme, 2000). The sequences of the hmw1A and hmw2A genes are identical for the first 1259 bp, thereafter diverging, with the HMW1 and HMW2 proteins exhibiting different cellular binding specificities, such that HMW1 binds to siaylated glycoproteins, whereas the receptor for HMW2 remains unknown (Giufre` et al., 2006). Diversity of the core-binding domains within HMW1 and HMW2 has been documented, although conserved sequence motifs have been identified. It is not known whether these core-binding domains elicit cross-reactive antibodies (Giufre` et al., 2006).

HMW1/HMW2 proteins are expressed by approximately 75 % of NTHI strains (Barenkamp & Leininger, 1992; St Geme et al., 1998), with the remaining 25 % of isolates that lack HMW1/HMW2 proteins expressing a second distinct class of adhesin known as Hia, an immunogenic high-molecular-mass protein recognized by human convalescent-phase serum antibodies and an allele of the Hib Hsf protein (Barenkamp & Bodor, 1990; Barenkamp & St Geme, 1996; Winter & Barenkamp, 2009). However, the presence of HMW1 and HMW2 among NTHI isolates may be more prevalent among invasive NTHI strains than those recovered from the middle ear and throat (Eccevit et al., 2004), indicating significant genetic variability among NTHI isolates.

Additional typing

An MLVA typing scheme has been devised for H. influenzae (Schouls et al., 2005). This four-locus scheme
Table 2. Typing approaches for *H. influenzae*

<table>
<thead>
<tr>
<th>Target for typing</th>
<th>Techniques available</th>
<th>Online database</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommended techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serotype</td>
<td>Slide agglutination, PCR, DNA sequencing</td>
<td>No</td>
<td>Falla et al. (1994)</td>
</tr>
<tr>
<td>Antibiotic susceptibility</td>
<td>PCR and DNA sequencing</td>
<td>No</td>
<td>Dabernat et al. (2002); Hasegawa et al. (2004); Ubukata et al. (2001)</td>
</tr>
<tr>
<td><strong>Additional typing techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OMP P2 and P5</td>
<td>PCR and sequencing</td>
<td>No</td>
<td>Mullins et al. (2009)</td>
</tr>
<tr>
<td>HMW1 and HMW2</td>
<td>PCR and sequencing</td>
<td>No</td>
<td>Giufre et al. (2006)</td>
</tr>
<tr>
<td>MLVA</td>
<td>PCR and sequencing</td>
<td>No</td>
<td>Schouls et al. (2005)</td>
</tr>
</tbody>
</table>

has been shown to have higher discriminatory power than MLST amongst serotype b isolates and may therefore be more useful in outbreak investigations. The technique has also been used to investigate clusters of invasive infections due to NTHI, and also to demonstrate vertical transmission from mother to neonate (M. Slack, unpublished observations).

**S. pneumoniae**

**Epidemiology and vaccination**

*S. pneumoniae* can produce a number of different clinical manifestations, ranging from less frequent invasive disease, presenting mainly as meningitis and septicaemia, to the more common but generally non-invasive conditions, such as pneumonia, sinusitis and otitis media. In 2005, WHO estimated that 1.6 million people died each year from pneumococcal disease, including up to 1 million children less than 5 years old, most of whom were living in developing countries (WHO, 2007). In Europe, the reported incidence of invasive pneumococcal disease ranged from 0.3 to 1.8 per 100 000 in 2006 (Pebody et al., 2006).

The polysaccharide capsule is the primary virulence factor of the pneumococcus, and more than 90 distinct serotypes have been described. Four vaccines were licensed in Europe at the time of writing, which include a variable number of capsular serotypes: the older 23-valent ‘plain’ pneumococcal polysaccharide vaccine (PPV) and the newer pneumococcal conjugate vaccines (PCV7, PCV10 and PCV13), which conjugate the pneumococcal polysaccharide to a protein carrier. PPV provides moderate protection against invasive disease due to 23 serotypes (1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F, 33F) in adults (Jackson et al., 2003), but is ineffective in young children, the age group most at risk of disease. PCVs are highly efficacious and provide longer-lasting immunity in children less than 5 years of age, although they protect against a smaller number of serotypes: PCV7 (serotypes 4, 6B, 9V, 14, 18C, 19F, 23F), PCV10 (PCV7 serotypes, plus 1, 5, 7F) and PCV13 (PCV7 serotypes, plus 1, 3, 5, 6A, 7F, 19A) [Centers for Disease Control and Prevention (CDC), 2010; Bryant et al., 2010; Poehling et al., 2004; Vesikari et al., 2009]. PCV7 is currently being phased out and replaced by PCV13, both of which are produced by Pfizer (formerly Wyeth Vaccines). PCV10 is produced by GlaxoSmithKine Biologicals, and has the added advantage of the conjugate protein (protein D) being from NTHI, thereby providing some immunological protection against this organism as well as *S. pneumoniae* (Prymula et al., 2006).

**Current typing approach**

**Pneumococcal serotype.** The Quellung reaction, first described in 1902, remains the gold standard for capsular serotyping (Austrian, 1976; Sørensen, 1993); however, the high cost of antisera combined with cross-reactions and discrepancies between serotypes, in addition to some isolates being non-typable, are major complications of this procedure. Furthermore, this technique is dependent on the isolation and growth of the organism, which, in some cases, e.g. among children with suspected meningitis or pulmonary empyema, is negative due to the administration of antibiotics prior to or shortly after presenting to hospital. Latex agglutination is an alternative serotyping method that is slightly cheaper and simpler to perform, but also requires the isolate to be cultured. Several PCR-based serotyping methods now exist, which avoid many of the challenges of conventional serotyping and could, in principle, be used for serotyping culture-negative samples. Correctly identifying non-typable pneumococci remains a challenge, however.

Capsule production in *S. pneumoniae* is controlled by the capsule polysaccharide synthesis (*cps*) gene cluster located between the genes *dexB* and *aliA*, which are not involved in capsule biosynthesis, and close to the PBP genes *pbp2x* (upstream of *dexB*) and *pbp1a* (downstream of *aliA*) (Garcia et al., 2000). Serotype 37 is an exception, as the *cps* locus is defective; in this case the serotype is determined by *tts*, the serotype 37 synthase gene (Llull et al., 1999). The *cps* gene cluster contains genes responsible for synthesis of the serotype-specific polysaccharide, including (except serotype 3) *wzy* (polysaccharide polymerase gene) and *wzx* (polysaccharide-anneal protein gene)
charide flippase gene). At the 5’ end of the cps gene cluster are four relatively conserved ORFs, cpsA (wzg)-cpsB (wzh)-cpsC (wzd)-cpsD (wze) (Jiang et al., 2001). In addition, heterogenic regions located between the 3’ end of cpsA and the 5’ end of cpsB exist between and within serotypes (Jiang et al., 2001).

The determination of the complete sequences of the cps loci from all of the 90 known pneumococcal serotypes at the time of the study enabled the design of a simple sequence-based scheme identifying pneumococcal serotypes by sequential multiplex PCR (Bentley et al., 2006; Pai et al., 2006). The primer selection and their arrangement were initially optimized based on the capsular serotype distribution found in the USA, but this method has been successfully used in serotyping isolates from Brazil and South Africa (Dias et al., 2007; Njanpop Lafourcade et al., 2010), with further optimization and redesign of the system allowing serotyping of South Asian isolates (Saha et al., 2008a). A protocol permitting the identification of pneumococcal serotypes rapidly and reliably from isolate sets as well as from clinical specimens has been established, and is available at http://www.cdc.gov/ncidod/biotech/ strep/pcr.htm. Furthermore, real-time PCR assays have been described that enable the detection and identification of S. pneumoniae in clinical settings. These assays show high sensitivity, enabling the enhanced detection of S. pneumoniae in cerebrospinal fluid and blood samples (Marchese et al., 2011; Moore et al., 2010).

MLST. MLST of S. pneumoniae is based on DNA sequencing of fragments of seven housekeeping loci (aroE, gdh, gki, recP, spi, xpt and ddl) (Enright & Spratt, 1998). The assignment of alleles at each locus is carried out using the MLST website http://spneumoniae.mlst.net. To date, over 13,000 S. pneumoniae isolates have been deposited in the MLST pneumococcal website, encompassing 6500 STs along with the provenance, geographical location and year of isolation (Table 3). Antibiotic resistance profiles are also available for most of the isolates. MLST is becoming the molecular typing method used by many reference laboratories. MLST has also been used to genotype culture-negative samples directly from the primary specimen, e.g. cerebrospinal and pleural fluids (Enright et al., 2000; Obando et al., 2008).

One of the most striking findings to be revealed by the use of MLST to genotype pneumococci is the strong association between ST and serotype, meaning that in general, one serotype is predominant within a particular ST, with relatively few well-recognized exceptions, and this association has been shown to be true among pneumococci from widely different geographical locations (Beall et al., 2006; Brueggemann et al., 2003; Enright & Spratt, 1998). This serotype–ST association has made it possible to detect unusual variants that have arisen in the USA since the vaccine was introduced in 2000, including the first vaccine escape recombinants (Beall et al., 2006; Brueggemann et al., 2007; Moore et al., 2008; Pai et al., 2005).

The genetic relationships between serotypable and non-typable pneumococci have been investigated using MLST and partial sequencing of the pneumolysin gene, ply (Table 3) (Hanage et al., 2005), which revealed three classes of non-typable isolates: (i) those with STs identical to those of serotypable pneumococci, suggesting downregulation or loss of capsule expression; (ii) those that clustered among serotypable pneumococci but which had different STs; and (iii) a more diverse set of isolates that did not cluster with serotypable pneumococci. Note that the ddl locus used in S. pneumoniae MLST typing is associated with a high frequency of interspecies recombination because of a physical linkage with resistance-conferring pbp2b, and consequently is not often used in phylogenetic analyses (Enright & Spratt, 1999).

Antibiotic susceptibility

Antimicrobial resistance in pneumococci may involve the modification of chromosomal genes, often by horizontal gene transfer of DNA that encodes altered proteins (e.g. PBPs) or the acquisition of chromosomally located transposons that encode resistance determinants (e.g. conferring tetracycline, kanamycin, erythromycin or chloramphenicol resistance) (Klugman, 1990). Measurement of antimicrobial resistance forms a major part of surveillance, and a number of national and international networks have been established to monitor this. Despite increasing levels of resistance, the two most widely used antimicrobial therapies used to treat pneumococcal infections remain the β-lactams and macrolides.

β-Lactam resistance. Resistance to β-lactams in S. pneumoniae is mediated by alterations in PBPs, of which modifications to PBP2x, PBP2b and PBP1a are the most important and result in their decreased affinity for β-lactams (Coffey et al., 1995; Grebe & Hakenbeck, 1996; Muñoz et al., 1992). PBP2x and PBP2b are primary resistance determinants with rearrangements conferring low-level resistance, whilst the presence of a low-affinity PBP1a is essential for high-level resistance, but requires a modified PBP2b and/or PBP2x.

Molecular typing of the PBPs can be obtained by amplifying the pbp1a, pbp2b and pbp2x genes. Amplified genes are then digested with HaeIII/DdeI (pbp1a) and HaeIII/Rsal (pbp2b and pbp2x) and electrophoresed on 3% gels (Gherardi et al., 2007). Molecular characterization of PBPs can also be performed by sequencing pbp1a, pbp2b and pbp2x (Nagai et al., 2002), but this method is less widespread. Sequencing of (at least) the DNA regions encoding the active sites and regions of nucleotide diversity requires multiple PCR and sequencing reactions per gene, and while this is highly useful for specific research questions, it is not particularly amenable to routine surveillance work. Furthermore, curated databases reliably linking sequence polymorphisms to susceptibility level, as for meningococci, are not available, because a wide
diversity of different nucleotide changes can be found individually and in combination among \( pbp \) genes in penicillin-non-susceptible pneumococci, making a reliable link(s) between nucleotide changes, MIC and clinical outcome difficult to identity. Nonetheless, a sequence database devoted to defined PBP gene segments associated with different \( \beta \)-lactam MICs may be a useful resource.

**Macrolide resistance.** Resistance to macrolide antibiotics such as clindamycin is mediated by two major mechanisms: methylation of ribosomal macrolide target sites, encoded by the gene \( erm(B) \), and drug efflux, encoded by the \( mef(A) \) gene (Farrell et al., 2005; Klugman, 1990, 2002). Erm(B)- or \( erm(B) + mef(A) \)-positive strains have high resistance levels and are resistant to clindamycin (MLS phenotype), whereas \( mef(A) \)-positive strains generally have lower resistance levels and are susceptible to this antibiotic (M phenotype). A frequent association of erythromycin and tetracycline resistance is often related to insertion of \( erm(B) \) into a conjugative transposon of the Tn916 family that harbours the \( tet(M) \) gene and carries integrase (\( int \)) and excisase (\( xis \)) genes. Members of this family that carry \( erm(B) \) include Tn6002, Tn1545 and Tn3872 (Bricianci et al., 2007). The two main subclasses of \( mef \) in \( S. pneumoniae \), \( mef(E) \) and \( mef(A) \), are carried on different but related elements: \( mef(A) \) on Tn1207.1 or Tn1207.3, and \( mef(E) \) on an element called ‘macrolide efflux genetic assembly’ (\( mega \)).

Antimicrobial resistance in European countries is monitored by a network of national surveillance systems known as the European Antimicrobial Resistance Surveillance Network (EARS-Net). It is the largest publicly funded surveillance system for antimicrobial resistance in Europe, with at present 900 public health laboratories serving over 1400 hospitals taking part. EARS-Net performs surveillance of antimicrobial susceptibility of seven bacterial pathogens commonly causing infections in humans, including \( S. pneumoniae \), the data for which can be accessed via http://www.ecdc.europa.eu/en/activities/surveillance/ears-net/pages/index.aspx (Table 3). EARS-Net surveillance has shown that resistance profiles of \( S. pneumoniae \) populations vary both geographically and temporally. For example, within Europe during the late 20th and early 21st centuries, penicillin non-susceptibility increased in Finland and Ireland, whilst France, Israel, Lithuania and Norway exhibited a decrease in such isolates. At the same time, erythromycin non-susceptibility rose in prevalence in Turkey and Ireland, while it decreased in six countries, including Belgium, France and the UK. Currently, across Europe, penicillin and erythromycin non-susceptibility are <1–37 and <1–23 %, respectively.

**Potential typing candidates: PspA and Ply**

**Pneumococcal surface protein A (PspA).** Additional molecular typing methods have been proposed for \( S. pneumoniae \) (Table 3). The virulence factor encoded by \( pspA \), pneumococcal surface protein A, is surface-exposed, highly immunogenic and has been shown to elicit protective antibodies against pneumococcal infection in animal models (McDaniel et al., 1998). There is continued interest in vaccines composed of this protein; however, consistent with its exposure to the immune system, it is one of the most polymorphic \( S. pneumoniae \) gene products. Six distinct \( pspA \) clades, classified into three families, have been identified (Hollingshead et al., 2000). Distribution of these families in pneumococcal populations varies, with over 90% of pneumococci expressing either family 1 or family 2 PspA (Melin et al., 2008). PspA-based vaccines may be an important alternative for protection against \( S. pneumoniae \), although the diversity of this gene will need to

### Table 3. Typing approaches for \( S. pneumoniae \)

<table>
<thead>
<tr>
<th>Target for typing</th>
<th>Techniques available</th>
<th>Online database</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommended techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serotype</td>
<td>Quellung reaction, latex agglutination, PCR</td>
<td>No, but methods detailed at <a href="http://www.cdc.gov/ncidod/">www.cdc.gov/ncidod/</a> biotech/strep/pcr.htm</td>
<td>Bentley et al. (2006); Dias et al. (2007); Njanpop Lafourcade et al. (2010)</td>
</tr>
<tr>
<td><strong>Antibiotic susceptibility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Additional typing techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pneumolysin, ( ply )</td>
<td>PCR</td>
<td>No</td>
<td>Ceyhan et al. (2010)</td>
</tr>
<tr>
<td>PFGE</td>
<td>Restriction digestion</td>
<td>No</td>
<td>Tenover et al. (2010)</td>
</tr>
<tr>
<td>MLVA</td>
<td>PCR</td>
<td>No</td>
<td>van Belkum et al. (2007)</td>
</tr>
</tbody>
</table>
be catalogued further in order to create a vaccine that provides maximum coverage.

**Pneumolysin (Ply), the autolysin gene lytA and Spn9802.** Amplification of the ply gene, encoding pneumolysin, has been used to detect *S. pneumoniae* in the laboratory confirmation of pneumococcal cases, and sequencing of this locus is often carried out alongside MLST (Ceyhan et al., 2010); however, detection of this gene is non-specific for the identification of *S. pneumoniae* in respiratory secretions as well as blood samples, and it is possible that false-positive results may be caused by alpha-haemolytic streptococci originating from normal oral flora that have entered the bloodstream (Abdelhaim et al., 2010). In addition, the commensal streptococcal species *Streptococcus mitis* and *Streptococcus oralis* have been reported to have the ply gene (Whatmore et al., 2000). Consequently, ply PCR is not considered suitable for routine detection of non-bacteraemic pneumococcal pneumonia, although, in combination with other conventional microbiological and/or nucleic acid-based methods, it may provide helpful information in the identification of the pneumococcus in certain circumstances. On the other hand, the presence of the autolysin gene lytA and Spn9802, a DNA fragment with an unknown function, is thought to be specific for *S. pneumoniae*, and has been successfully used to identify the organism in plasma samples from patients with community-acquired pneumonia by real-time PCR. These genes have potential application in the rapid diagnosis of bacteraemic pneumococcal pneumonia, but this needs to be extensively validated before the approach can be routinely used in a diagnostic setting (Abdelhaim et al., 2010).

**Additional typing**

**PFGE.** PFGE is a genotyping technique that has been used for several decades and is still used in many laboratories (Tenover et al., 2010). PFGE has several advantages as a genotyping method: (i) it is a genome-wide analysis, albeit in fragments visualized on an agarose gel; (ii) it is a high-resolution typing method and standardized protocols exist; (iii) DNA restriction patterns are stable and reproducible; and (iv) it has been established in a number of reference laboratories and has a cost perceived as inexpensive. However, there are also many disadvantages in using PFGE as a genotyping tool. Intra- and inter-laboratory variability is a major problem, while bands on a gel can be difficult to interpret and data are not easily portable. In addition, the nature of the genetic variation being indexed is poorly understood; there is the question of whether bands of the same size in two isolates really represent the same pieces of DNA, and there is no consensus naming scheme. Based on the above advantages and disadvantages, MLST is viewed as the preferred pneumococcal genotyping technique, and as the cost of sequencing consumables continues to decrease, sequence-based methods such as MLST will become affordable for routine reference laboratories.

**MLVA.** MLVA is a PCR-based typing method that indexes the variability encountered in regions of repetitive DNA (van Belkum et al., 2007). Repetitive DNA is often incorrectly copied in bacterial species, resulting in shortening or lengthening of the repeat regions. The resultant variation can be assessed by performing repeat-spanning PCR amplification and determining the length of the PCR product, or by sequence extension reactions. The technique has been applied to *S. pneumoniae*, successfully identifying a cluster of invasive pneumococcal disease due to serotype 5 *S. pneumoniae* (Sadowsy et al., 2010). This approach has not been routinely applied to pneumococci, and suffers from the major drawback that the evolution and rate of change of repetitive DNA are so rapid that it is unsuited to establishing the relationships of anything but the most closely related strains (Lindstedt, 2005).

**Conclusions**

The last decade of the 20th century and the first decade of the 21st have seen the increasing application of molecular, and especially DNA-based, typing to the characterization of invasive encapsulated bacteria. These methods have complemented, and have the potential to replace, the phenotypic methods that dominated the characterization of these bacteria for much of the 20th century. Molecular approaches have a number of advantages, including rapidity, reproducibility, lack of reliance on specific reagents, and portability among laboratories. They can also be applied to clinical specimens, from which it is not always possible to obtain a bacteriological isolate. While there remain some issues of training, infrastructure and perceived cost, it is almost certain that the coming decade will see the replacement of many conventional phenotypic tests currently performed in clinical and reference laboratories with some form of molecular assay.

We recommend that, as much as possible, these assays should be those based on nucleotide sequence determination of key genomic loci. Generic technology can be used across all pathogens, so that laboratories can become generalists, equally adept at characterizing widely different pathogens. This is not, however, to say that the phenotype is no longer relevant. On the contrary, the phenotype remains the crucial point at issue for the clinical microbiologists and epidemiologists who need to know whether an isolate is, for example, a vaccine escape variant, poses particular risks because of its antibiotic resistance profile, or is especially virulent or likely to cause a disease outbreak.

The success of molecular typing over the coming decade will therefore depend on the linkage of sequence data to relevant phenotypic data, so that sequences can be used reliably to predict the clinical properties of the bacterial isolates characterized. This will require the international
availability of assured quality schemes and the free exchange of relevant and appropriate data, which will be facilitated by the internet. The PubMedLST Neisseria website provides a paradigm for this type of resource. It is freely available, managed by an international group of scientists and clinicians active in the meningococcal community, and includes both genotypic and phenotypic data. The reference set of alleles for various loci makes it possible to readily identify from sequence alone key properties, such as membership of a particular serogroup or hyperinvasive lineage, or susceptibility to clinically important antibiotics.

If recent, rapid developments in nucleotide-sequencing technologies, especially the so-called ‘Next-Gen’ sequencing methods, continue, it is inevitable that sequences will become ever cheaper and more easy to determine. The challenge, therefore, is to correlate this increasing resource of sequence information with phenotypic properties of clinical importance and to make this information available to the community internationally, thereby ensuring that these advances are effectively translated into clinical benefits.

Acknowledgements

This work was funded by the ECDC-funded Invasive Bacterial Disease Laboratory Network (IBD-LabNet).

References


Molecular typing methods for bacterial infections


